NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

# Unique identifiers for training materials

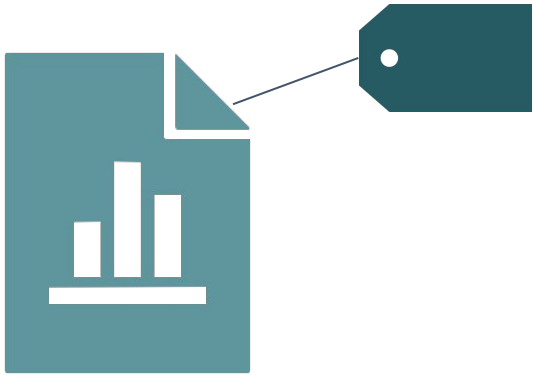
Elin Kronander

# Learning outcomes

By the end of this session, you should be able to:

- **Explain** what unique persistent identifiers are and their benefits
- **List** and differentiate the types of unique identifiers that are relevant for publishing and sharing training materials
- **Compare** different strategies for unique identifiers for training materials
- **Create** versioned DOIs for training materials

# What is a unique persistent identifier (PID)



Identify

Verify

Locate


JANE
JANE DOE, MILL RD 1
SWEDEN
JANE DOE, HIGH ST 5
NORWAY
JANE **SMITH**, CASHEW WAY
GERMANY

# Features of PIDs

- Globally unique
  - It should comply with a **controlled syntax** to avoid clashes
- Persistent
  - It should be **maintained** for a **long period of time.** The syntax used for the identifier should also be persistent
- Resolvable
  - It should allow both **human** and **machine** users to **access the resource**

# Digital Object Identifier - DOI

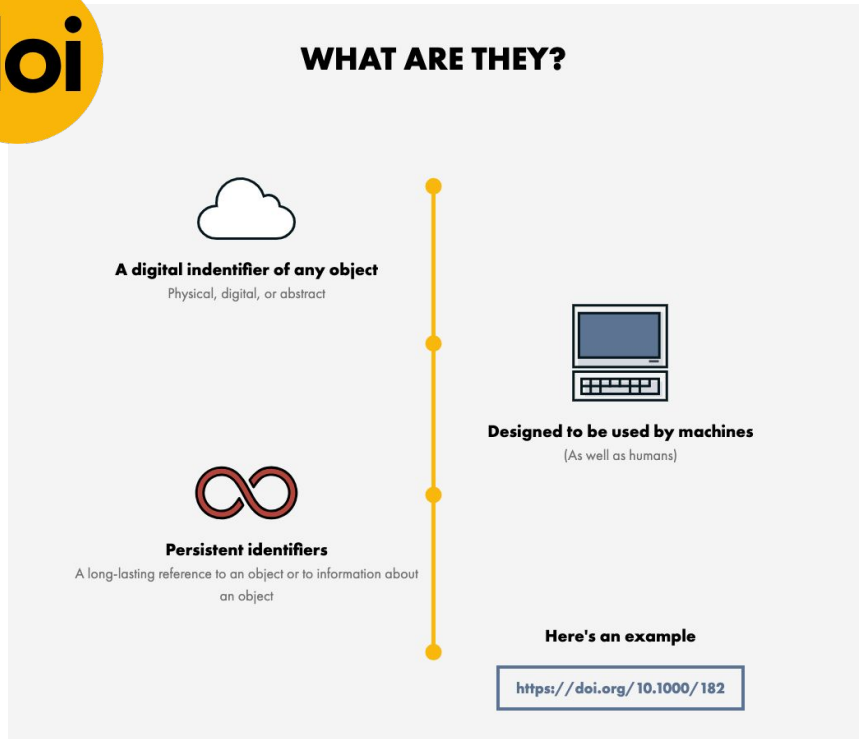



Illustration from the DOI Foundations website
https://www.doi.org/the-identifier/what-is-a-doi/

# Digital Object Identifier - DOI

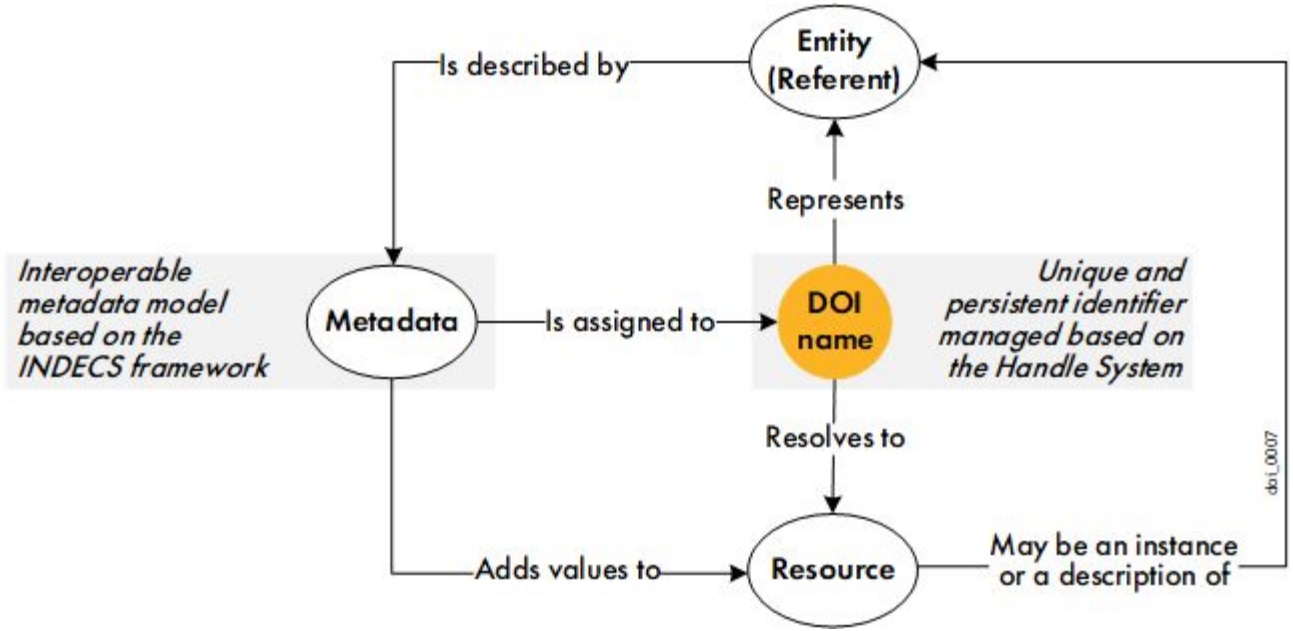



Illustration from The DOI Handbook April 2023 (https://doi.org/10.1000/182 identifies the latest current version of the handbook)

# Other relevant PIDs

ORCID - Open Researcher and Contributor ID

- persistent identifiers for researchers
- takes homonymy into account
- add aliases to your profile if your name changes
- ORCID stays the same when affiliation changes

The Research Organization Registry

- persistent identifiers for research organizations

# Benefits of PIDs

- uniquely distinguish resources from similar objects (F)
- a place to keep the metadata (F)
- machine actionable identifiers increase findability (F)
- resolves providing a way or information on how to access the object (A)
- enhances citability leading to easier reuse (R)

# Reflection

In the context of training materials why are PIDs needed?

Which identifier should be used for each need?

PIDs can help distinguish between:

- different materials - **DOI**
- different versions of the same material - **DOI**
- different authors and contributors - **ORCID**
- different organisations - **ROR**

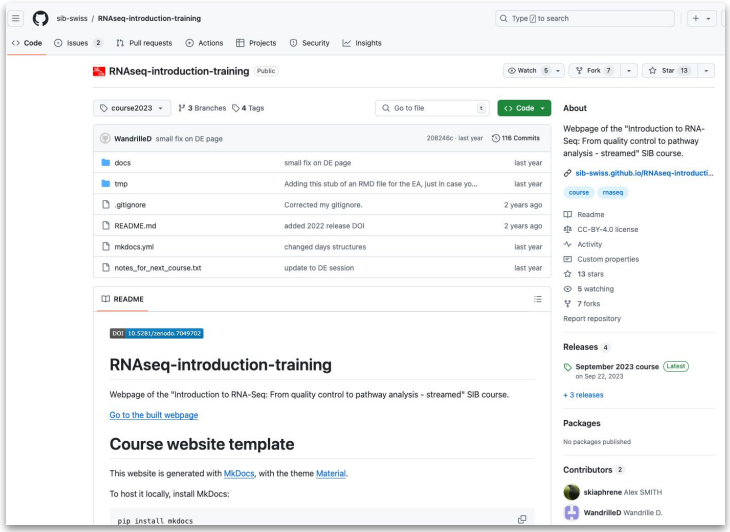
# Case Study

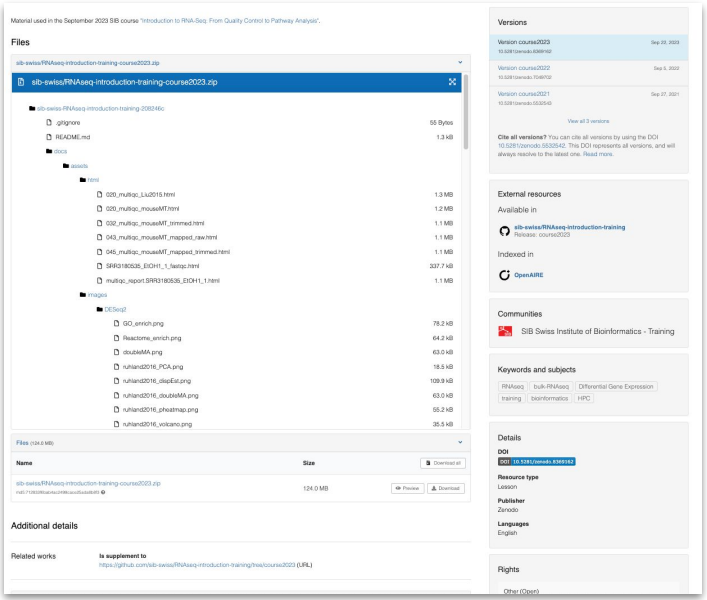
1. Group discussion 10 minutes
    a. Assign 1 person to take notes in the shared document
    b. Read through the use case assigned to your group from the FAIR training handbook
    c. Discuss and write down a short summary of the strategy used as well as pros and cons with this strategy

2. Plenary discussion 10 minutes
    a. Each group will share their observations and reflections

# GitHub integrations

GitHub ⟷ zenodo

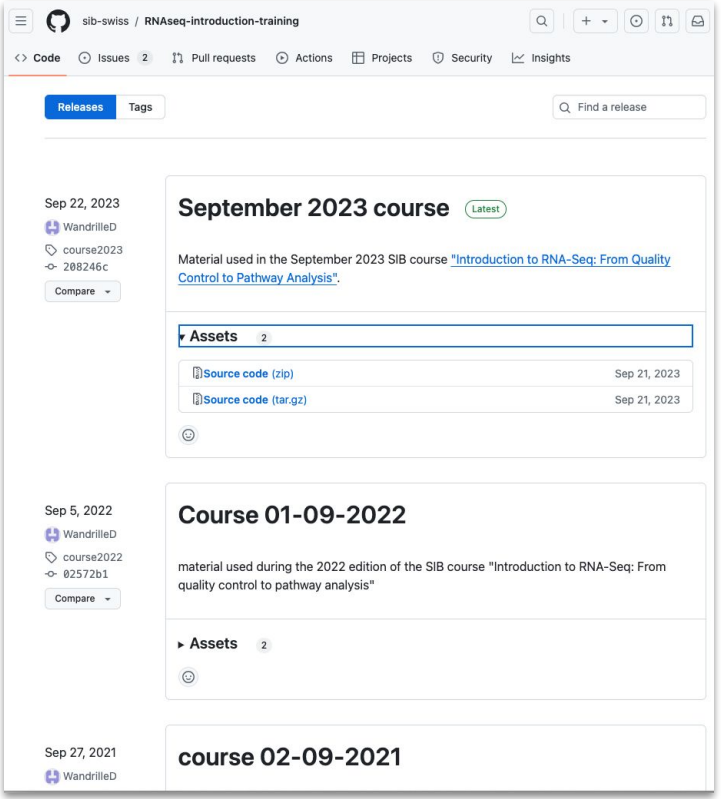


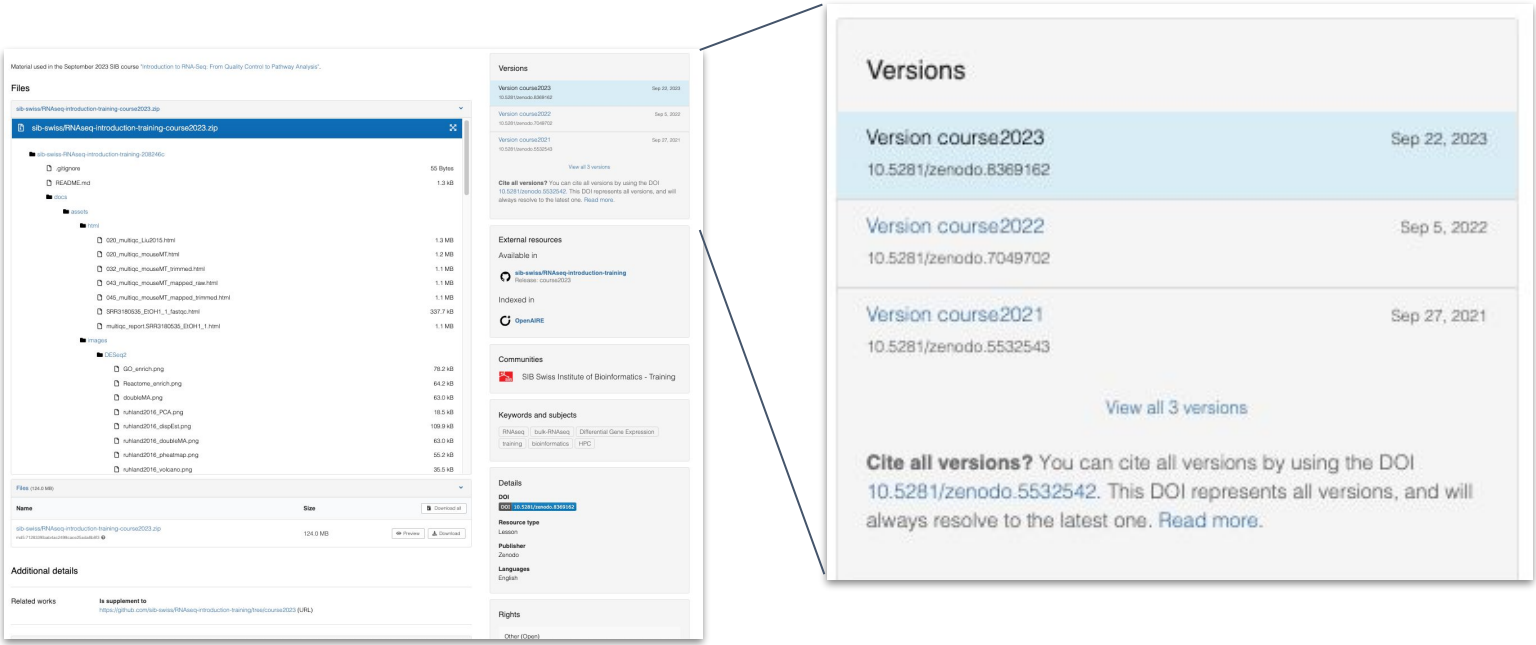

doi 10.5281/zenodo.5532542

# GitHub releases

- Snapshot of project at specific time point
- Packaged with re-use in mind
- Downloadable (zip file and tarball)
- Attached with version number/name via a tag
- Release notes to describe the specifics of the snapshot



# Versioned DOIs with releases



# Reflection

How would a good strategy for your own context look like?

Things to consider

- Do you want to get a PID for each training material?
- Do you want to get one PID for your whole training or course?
- Do you want to get a separate PID for each topic/module? For example, for a course containing several topics.
- Do you want to create a collection of topics with a PID where each concept will also have a PID and associated metadata?
- Do you want to get one PID for your whole training or course?

# Introduction to Tutorial

1. Use zenodo Sandbox to get a DOI for your training material
    a. By using GitHub integration
    b. By manually uploading a zip file of your GDrive folder
2. Enrich the Zenodo record with the metadata from previous session
3. Add the DOI to your hosting platform
4. Add the DOI to your TeSS record

Go to the tutorial in Chapter 08